

GENESIS: The General Earth Science Investigation Suite



Tom Yunck, Brian Wilson, Amy Braverman, Elaine Dobinson and Eric Fetzer

Jet Propulsion Laboratory

“Turning the accomplishment of many years into an hour-glass”

-- Henry V (1,i)





The Vision of Earth System Science



Characterize
Earth's varied
behavior

Understand
the Earth as an
integrated system

Predict
Earth's
response to
complex forcings



Today's Earth Science IT Challenges

- Coping with diverse Earth science data sets:
 - **Locating** the right products (**Data Discovery**)
 - **Selecting**: browse, query, subset, customize...
 - **Retrieving** large data volumes swiftly
 - **Fusing** diverse, incommensurate products
 - **Visualizing** massive multidimensional data
 - **Discovering** knowledge: Summarize/Analyze/Mine/...
 - **Predicting**: Data Assimilation, Earth System Modeling: Tools / Environments / Frameworks / CPU
- Sample research scenario Today: Multi-year effort for a modest, cross-instrument study



Computing Paradigms

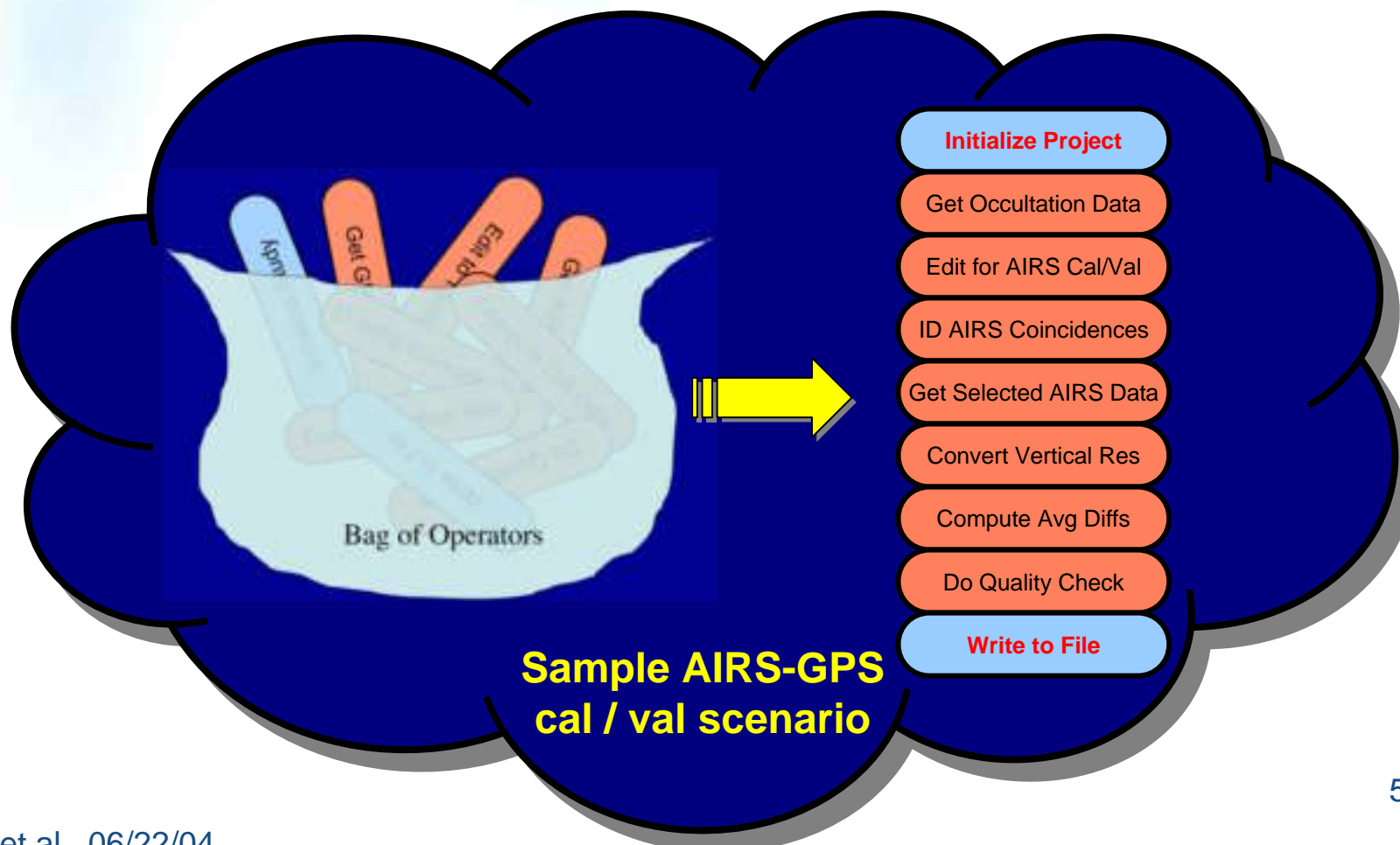
- **Old:** Big Iron mainframe with many users
- **Current:** Desktop work stations connected to Internet;
- **Emerging:** The Global Grid – Computing as a utility
 - Desktops, supercomputers, storage, tools joined as one
 - Petaflops of cpu, petabytes of storage
 - Bulk bandwidths: hundreds of GB/sec
 - Secure, services-based architecture
 - Vast library of analysis and modeling tools
 - Real time 3D visualizations, animations
 - Semantic understanding of service requests
 - Global-scale computing on your desktop





Vision Research Scenario

- **Visual Programming:** Drag-and-drop icons to create end-to-end research flow in ~hours rather than years



iEarth



- New Project... ⌘N
- Open Project... ⌘O
- Open Recent ▶
- Save Project ⌘S
- Save Frame As... ⌘F
- Import... ⌘⌘I
- Export... ⌘⌘E
- Show Info ⌘I
- Empty Trash...

Layout



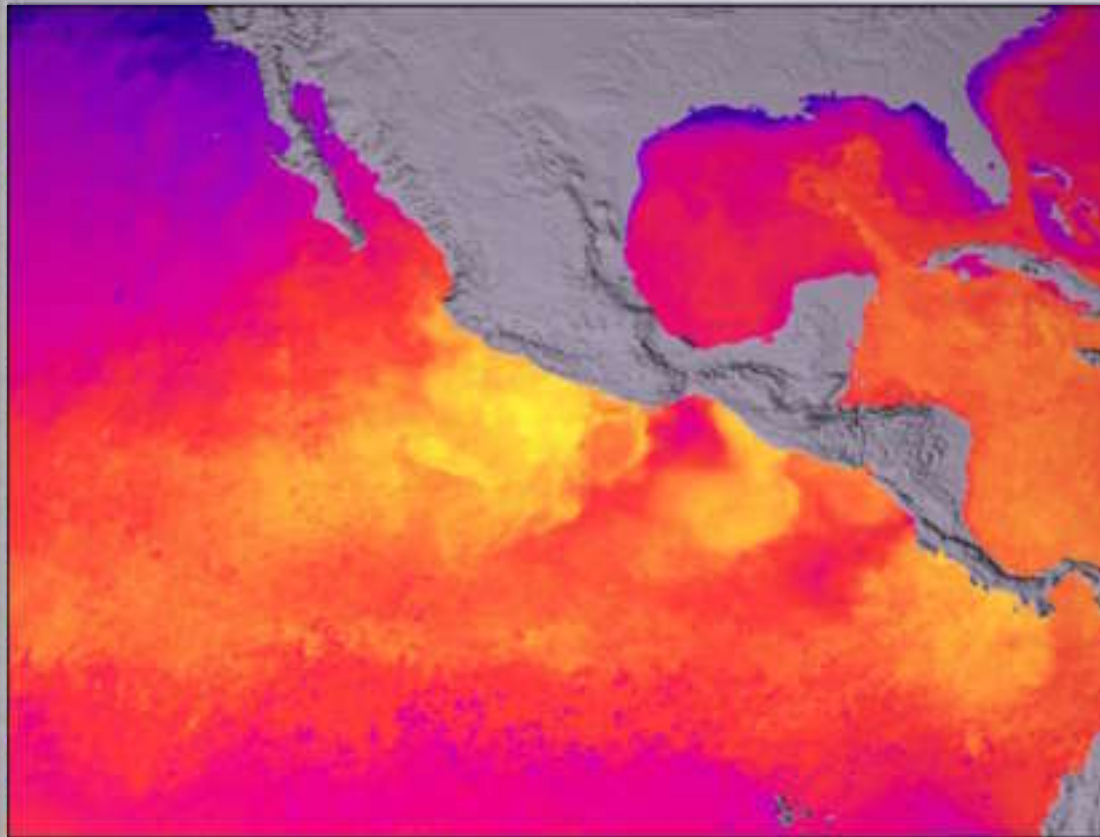
Setup Measurements Customize Summarize Visualize Analyze Archive

Drag clips here to build your project.

0 KB

8.23 GB free

iEarth Setup



Visualize Analyze Archive

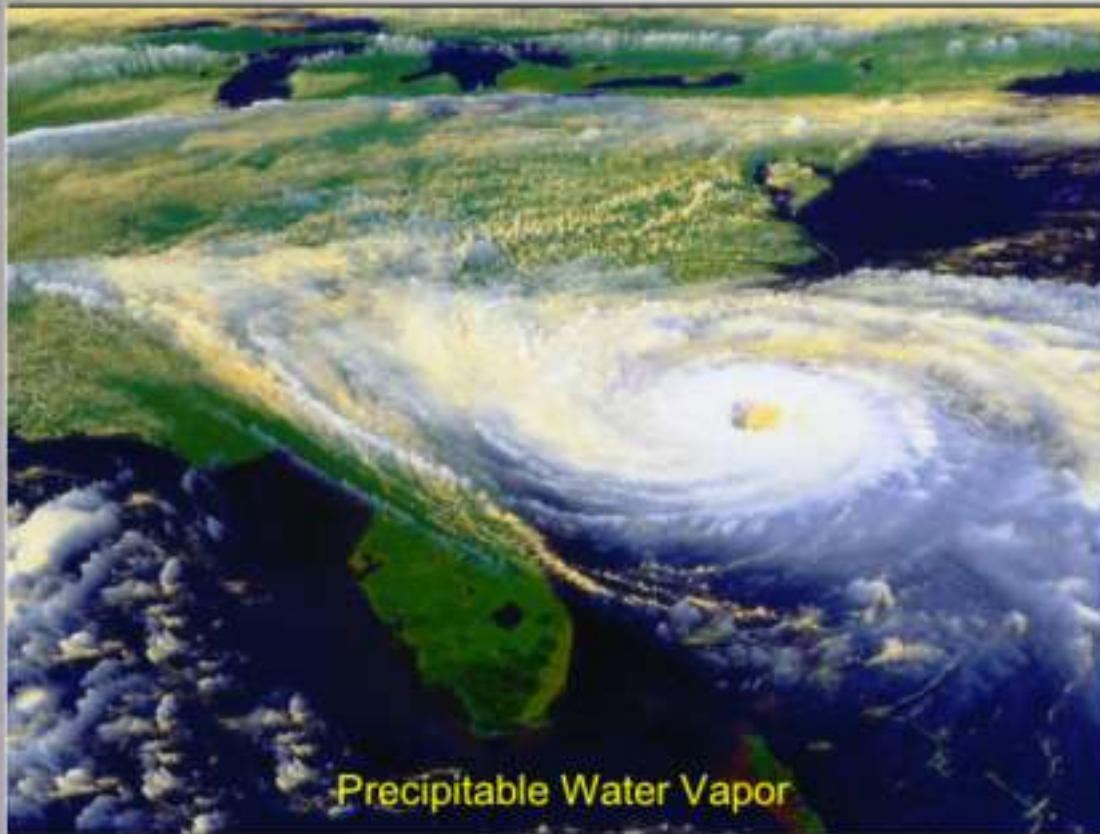


Drag clips here to build your project.



The **NASA** Earth Measurement Set

iEarth Measurements

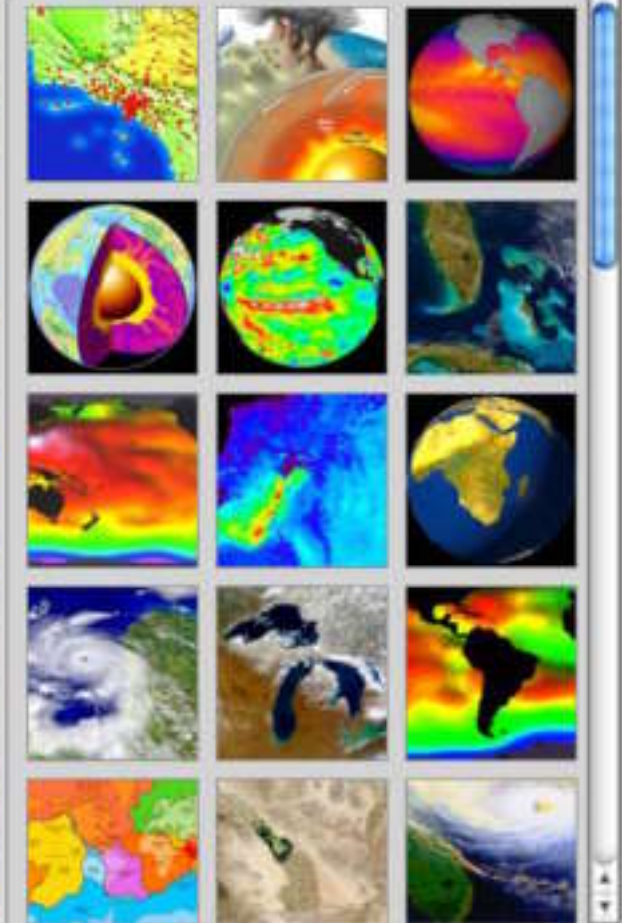


00:00

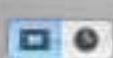


iEarth Visualize

Operators



00:00



Setup

Measurements

Customize

Summary

Visualize

Analyze

Archive

General Earth Science Investigation Suite

GENESIS

Powered by **SciFlo™**

0 KB

8.23 GB free



Under the Hood: Five Core Ideas

- Loosely-coupled distributed computing using SOAP
- Exposing scientific analysis operators as SOAP web services
- Specifying a processing stream as an XML document
- Dataflow engine for a parallel execution and load balancing
- Visual point-and-click programming



Guiding Themes

■ Decentralization

■ The Global Grid

■ Peer-to-Peer

■ Machine-to-machine

■ Automated workflows

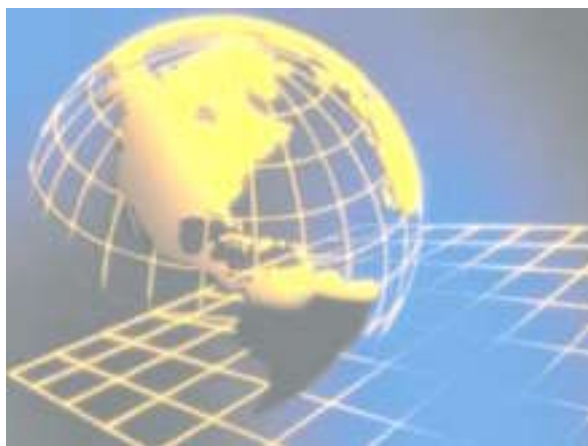
■ Distributed execution

■ Dynamic load balancing

■ Grid web services

■ Multi-scale integration

■ Plug-and-play software





Data Discovery...

■ Current:

- GCMD, EDG, ECHO, FIND, SRB, UDDI, ...
- Integrated archives: EOS-Webster (UNH), GLCF (UMD), TRFIC (Mich State), GHRC (Huntsville), and many others
- Co-op and commercial services: UNIDATA; Coop Climate Rainfall Data Center (CSU); Earth Data Discovery Consortium

■ Emerging:

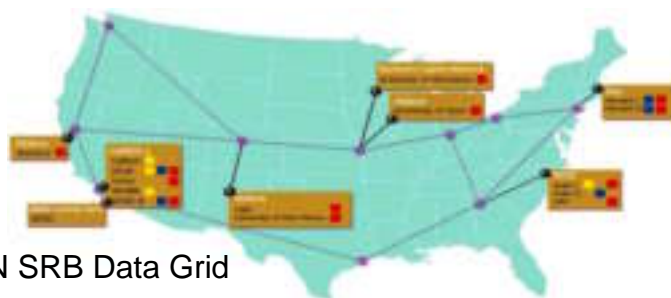
- Persistent archives and logical namespaces
- Proliferating peer-to-peer exchange networks
- Content-based-search / Semantic Web: Semantic representation of data enabling computers to understand web content.
- SRB+, NASA Intelligent Archive Program,...
- Unidata THREDDS thematic data registries





Grid Technology, Peer-to-Peer Model

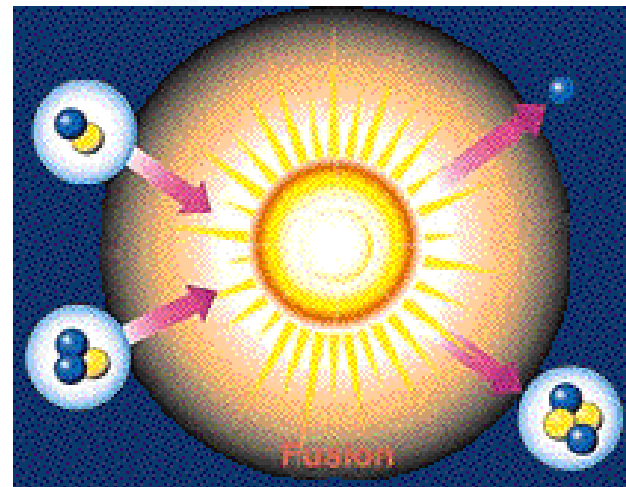
- Flat, decentralized storage, discovery, distribution
- Multicasting / swarming: proliferating indexed local data sets
- Secure electronic data verification / pedigree
- XML messaging among peers
- Scalable queries, access, replication





Controlled Data Fusion

- Optimal combining of like quantities
 - Fitting NASA's "Missions-to-Measurements" theme
 - Bayesian hierarchical modeling and optimal estimation
- Re-projection and co-registration
 - ESMF re-gridding utilities
 - OGC: WMS/WCS – Geo-referencing info in metadata
 - Image co-registration tool for NGA (JPL)
- ESIP Federation Activities
 - GENESIS: Data Fusion for Multi-Sensor ESS (JPL)
 - Synthesizing Carbon, Water & Energy Cycle Prods (GSFC)
 - Inter-Service Data Integration for Geodetic Operations (JPL)

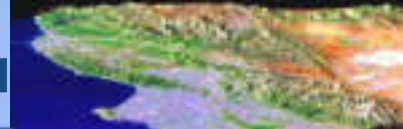




Standards & Interoperability

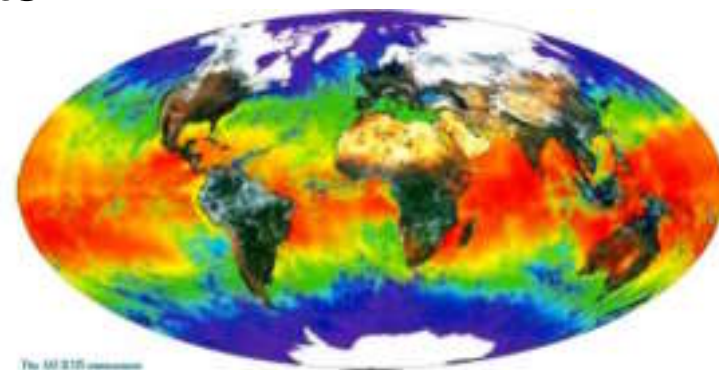
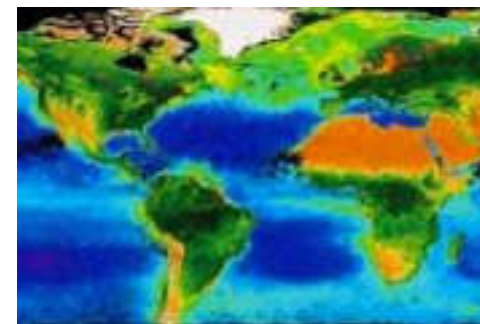
- **Grid Services:** Web Services Resource Framework (WSRF), Globus Toolkit, National Middleware Initiative (NMI), SRB
- **Logical namespaces:** Persistent URLs for data & services
- **Distributed control:** XML messaging; SOAP/WSDL/UDDI
- **Semantic Web:** Ontology Web Language (OWL)
- **Data and metadata:** content / formats / interfaces
 - OGC Standards: WMS, WCS, WFS, GML,...
 - Earth Science Markup Language (ESML)
 - FGDC / GCMD DIF / Dublin Core
 - OPeNDAP / NetCDF / HDF-EOS
- **Key standards activities:**
 - Open GIS Consortium
 - Global Grid Forum
 - OASIS, W3C
 - NASA Data System Working Groups





Summary: Directions and Themes

- Decentralized multiscale computing as a utility
- Peer-to-peer discovery, distribution, services
- Ubiquitous machine-machine communications
- Distributed control by XML messaging
- Semantic information representation
- Integrated modeling environments
- Visual workflow programming
- High-performance computing
- Standards Standards Standards



The 50 000 kilometers



Topics

- Data discovery and selection
- Grid technology, P2P model
- Mass data transfer
- Data visualization
- Data fusion and summarization
- Assisted knowledge discovery
- Environmental prediction / high-end computing
- Automated workflow / resource optimization
- Standards and interoperability





...and Selection (Browse, Query, Subset, Customize)

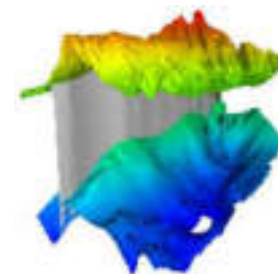
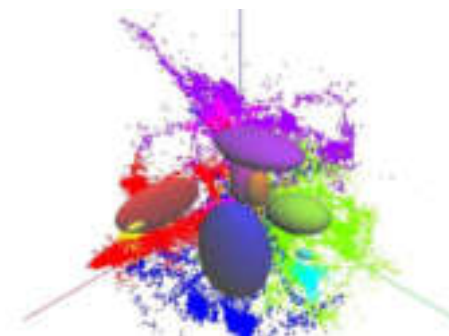
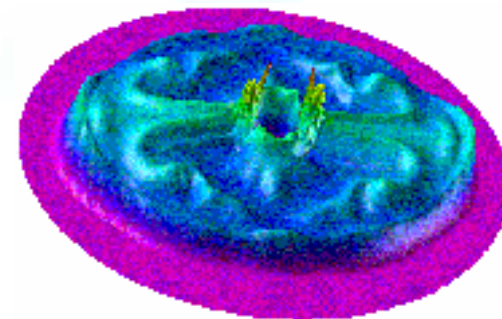
- Open GIS Consortium: Web Coverage Server
- Commercial DBMSs: Informix Geodetic Data Blade
- Subset.org: Portal to a wide variety of subsetting tools
- Data Interchange: ESML – Earth Science Markup Language (UAH) for data/tool interoperability and links to semantic web ontologies
- Federation:
 - Earth System Science Workbench (UCSB)
 - Mobilization of NASA EOS Data & Information Through Web Services (GMU)
 - Custom Order Processing (MSFC/UAH/RSS)
 - UNITE – Interchange Technology Prototype (UAH/JPL/ORNL)
 - Ongoing DAAC developments





Data Visualization

- Open GIS Consortium: WMS/WCS
- Some visualization tools:
 - Mapserver (WMS/WCS) (UMinn)
 - Viewer.digitalearth.gov (GSFC)
 - Live Access Server (DODS)
 - ImageTours (Wegman), Limn (Cook)
 - OurOcean, YourSky, Digital Light Table (JPL)
- ESIP Federation:
 - Immersive Earth (Rice U)
 - Virtual Interactive Environmental Worlds (US Sat Lab)
 - Visualization for Meso-American Biological Corridor and Beyond (MSFC, UAH)
 - Visualization of Yellowstone (CSU Monterey Bay)
- Commercial: Pixar, ILM, Magic Globe, Stormcenter





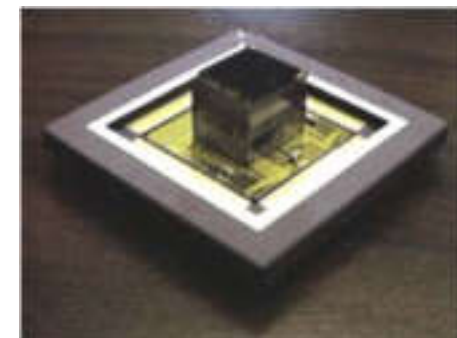
Assisted Knowledge Discovery

■ Advanced algorithms

- Feature extraction and clustering
- Data mining, aggregation, summarization
- Supervised / unsupervised machine learning
- Support vector machines for classification



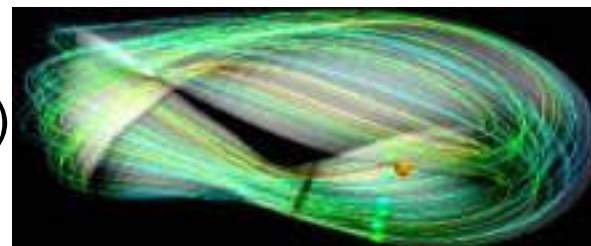
■ Optical correlators and hardware neural nets for rapid image analysis (JPL)



■ Complex data mining and analysis (UAH)

- Multi-level mining (ADaM)
- Mining in distributed/federated/parallel processing environments
- On-board and Event-driven mining (EVE)

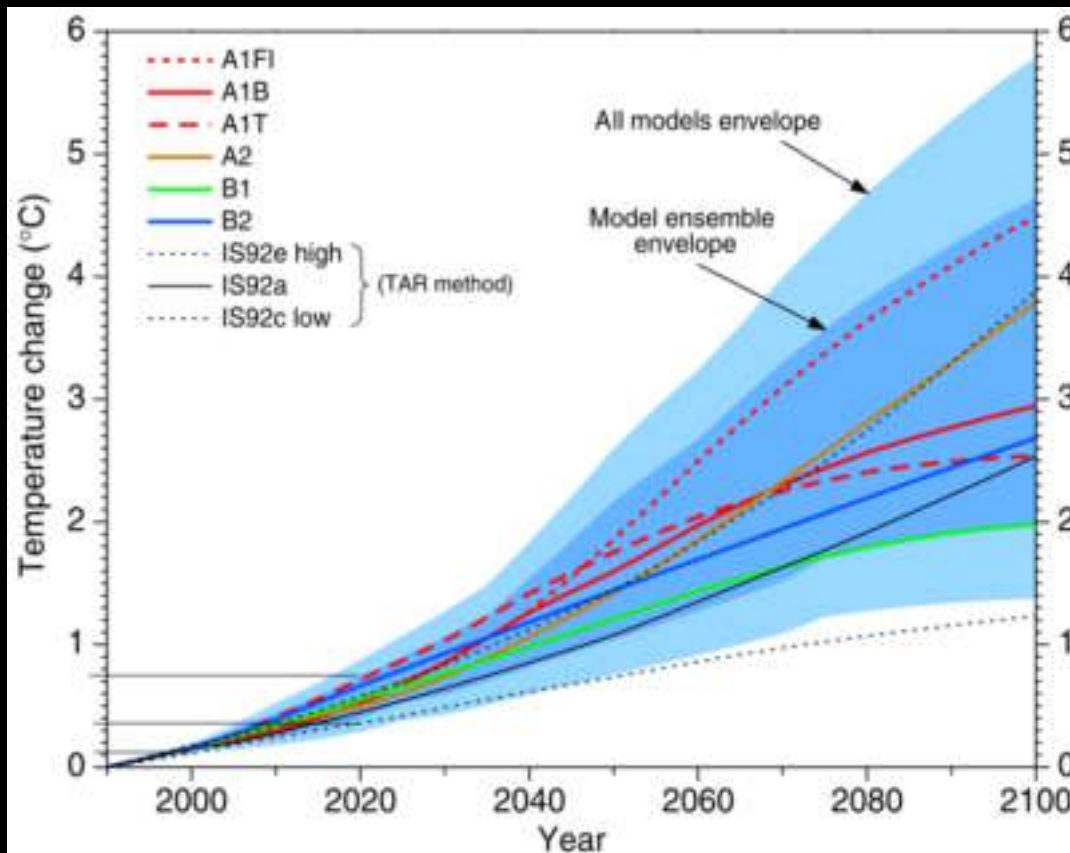
■ Highly iterated interactive analysis (GENESIS/SciFlo)





ESE Space Missions

Predicted Global Warming



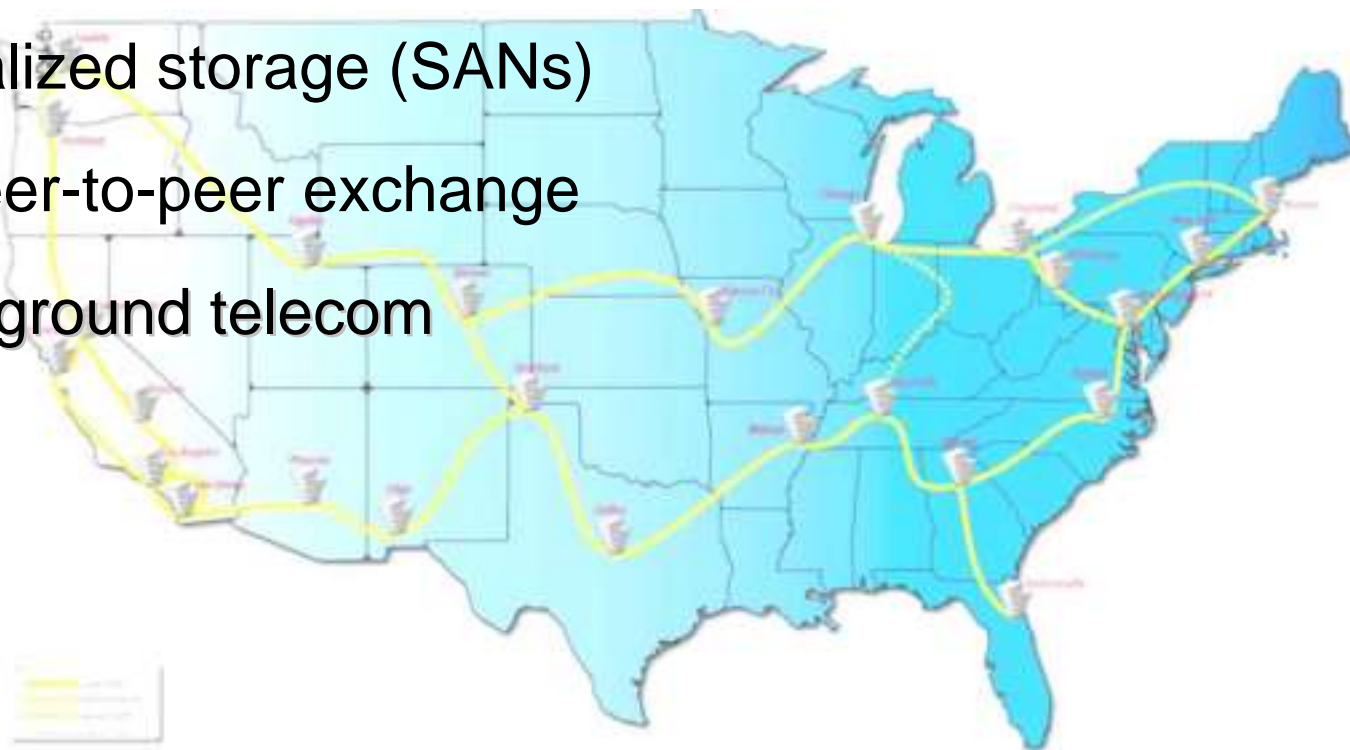
IPCC 2001





Data Mass Transit

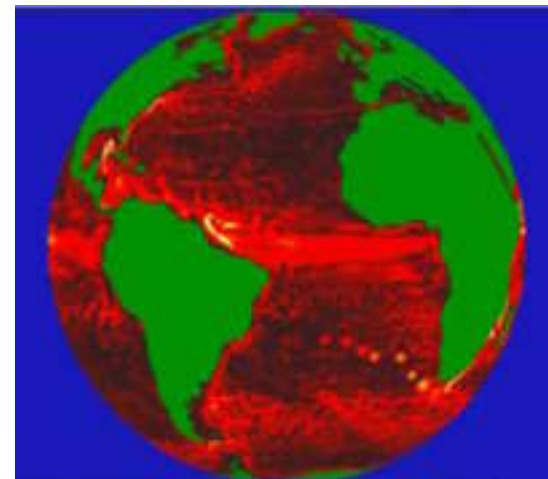
- Dedicated optical fiber – National LambdaRail
- Multicasting – Data pushed to many sites
- Decentralized storage (SANs)
- Direct peer-to-peer exchange
- IP air-to-ground telecom





Environmental Prediction

- Readily coupled, mixed-scale Earth system models
- Integrated modeling environments
 - Earth System Modeling Framework
 - Grid Compute Engine (AIST)
 - Data Assimilation SERVO Grid (AIST)
 - Linked Environments for Atmospheric Discovery (OU/UIUC/UAH/IU/UCAR)
 - Common component architecture
- High-performance computing
 - Grid Computing: TeraGrid, IPG,...
 - Earth Modeling supercomputer (Japan)
 - Terascale Computing Facility (Virginia Tech)
 - ECCO Simulation on the first 512 Node SGI Altix
 - DARPA: High-Productivity Comp Systems (Cascade)



ECCO Ocean Simulation

